[Communicative & Integrative Biology 2:5, 408-410; September/October 2009]; ©2009 Landes Bioscience

Article Addendum

Trans-species polymorphism, HLA-disease associations and the evolution of the MHC

Cock van Oosterhout

University of Hull; Hull, UK

Key words: major histocompatibility complex (MHC), mutational load, transposable elements, retrovirus, trans-species polymorphism, human leukocyte antigen (HLA), HLA disease-association, linkage disequilibrium, balancing selection, coevolutionary arms race

Currently, the paradigm is that major histocompatibility complex (MHC) polymorphism is maintained by balancing selection on the immune genes. However, other evolutionary forces besides selection also play a role in the population genetics of this multigene family. van Oosterhout proposed a new theory of MHC evolution called associative balancing complex (ABC) evolution.[1] This theory incorporates the effects of the evolutionary forces in the entire MHC region (peri-MHC), and it proposes that recessive deleterious mutations can accumulate in the peri-MHC in a process similar to Muller's ratchet.[2] These mutations are not easily purged because epistasis and high gene diversity in the MHC reduce the efficacy of natural selection. Because natural selection is less efficient, it could also make the MHC prone to the onslaught of genomic parasites such as retroviruses and transposable elements (TEs). The accumulated genetic load has important consequences for the evolution of this immune gene family, and it can reinforce linkage disequilibria and help to maintain the MHC polymorphism. ABC evolution offers new insights into some of the most puzzling aspects of the MHC, including the occurrence of identical MHC sequences in diverged species (i.e., trans-species polymorphism). It may also explain why the large numbers of disease-associated mutations are not removed by natural selection, and why the genes that protect vertebrates against infectious diseases are associated to such a wide variety of genetic disorders.

MHC genes are among the most polymorphic genes in the vertebrate genome. Several models of balancing selection have been proposed to explain this variation, including overdominance (heterozygote superiority),[3] negative frequency dependent selection (rare allele advantage),[4] and selection varying in time and space.[5] These theories were developed to explain how genetic polymorphisms can be maintained in populations, sometimes over extraordinary lengths of time. Fluctuations in allele frequencies due to random genetic drift are reduced by balancing selection, which increases the MHC gene diversity. The negative frequency dependent model of balancing selection proposes that rare alleles confer a fitness advantage, which reduces the likelihood of them being lost from the population.[4] Similarly, a high level of MHC polymorphisms is maintained when heterozygote individuals have a superior fitness, for example because their immune system can recognize a wider range of parasites than homozygotes.[3]

These models have been developed to understand the population genetics of single genes, but the MHC is a multigene family associated with non-neutral polymorphisms. Studies on the human MHC (the human leukocyte antigen, HLA) show that over 100 diseases are associated to these genes,[6,7] which implies that selection is not only acting on parasite recognition. This suggests that these simple models of balancing selection do not adequately capture all processes that shape the evolution and population genetics of this multigene family. Recently, van Oosterhout highlighted a number of paradoxes of the MHC that cannot be explained by these simple models of balancing selection and proposed a new theory of MHC evolution.[1]

The Paradox of Trans-Species Polymorphism

One of the most remarkable characteristics of the MHC is a phenomenon called trans-species polymorphism (TSP).[8] Identical MHC sequences can be found in diverged species that have been separated by millions of years of independent evolution. This is remarkable, given that these genes are involved in a coevolutionary arms race with parasites. Red Queen dynamics suggests that these genes should evolve rapidly,[9] rather than showing a low rate of nucleotide substitutions. The apparent lack in nucleotide substitutions gives them the appearance of being in an "evolutionary stasis", quite the opposite of what is expected

Correspondence to: Cock van Oosterhout; University of Hull; Hull HU6 7RX UK; Tel.: +44.1482.465505; Fax: +44.1482.465458; Email: C.Van-Oosterhout@hull.ac.uk

Submitted: 04/14/09; Accepted: 04/15/09

Previously published online as a *Communicative & Integrative Biology* E-publication:
http://www.landesbioscience.com/journals/cib/article/8765

Addendum to: van Oosterhout C. A new theory of MHC evolution: beyond selection on the immune genes. Proc R Soc B 2009; 276:657–65; PMID: 18986972; DOI: 10.1098/rspb.2008.1299.

for genes that should keep up with the genetic adaptations of fast evolving parasites.

This issue is furthermore confounded by the fact that several theoretical studies have shown that balancing selection could maintain such TSP.[10] However, the genealogies of MHC genes do not only show extremely long terminal branches (i.e., evidence of TSP), but also that the extant alleles are highly diverged from each other. Simple models of balancing selection cannot explain how diverged sequences can persist over such extraordinary lengths of time.[1,11] Once a population is in a mutation-selection-drift equilibrium, balancing selection will only increase the turnover rate of alleles. For example, with negative frequency dependent selection (i.e., rare allele advantage), novel mutations will be initially rare, and hence, they will offer a fitness advantage. Compared to neutral model of evolution, these selectively favorable new mutants are more likely to increase in frequency and become established in the population, thereby pushing out one of the older extant alleles.[1]

Why are There so Many Diseases Associated to the MHC?

In humans, the MHC multigene family is known as the human leukocyte antigen (HLA), and it occupies a relative small area (4 Mb) of our genome. However, the circa 230 MHC loci are collectively associated with more than 100 pathologies.[6] Ironically, the genes that have evolved to protect the host against infectious diseases are cause of many genetic disorders themselves. This poses the question: why can natural selection not remove these disease-causing mutations?

Some mutational effects may only become expressed late in life, and natural selection is less efficient removing those mutations because fewer individuals reach such old age, and those that do may already have reproduced. Such mutations find themselves in a "selection shadow" and may persist despite their deleterious effects.[12] It has also been suggested that the mutations that are in the current environment may lead to disease, but that they are not unconditionally deleterious. In some environmental conditions, such mutations may confer a fitness advantage through pleiotropic effects, or the polymorphisms are maintained by epistatic gene interactions. For example, the extended 8.1 HLA haplotype in humans appears to carry two loci with antagonistic effects.[13] Individuals carrying both loci showed autoantibody titers that were identical to healthy control individuals that did not carry this haplotype. However, recombinant 8.1 HLA haplotypes showed higher titers than baseline, which is typical for patients with thymus hyperplasia.[13] Apparently, the detrimental fitness effects are cancelled out when both mutations are present within the same extended haplotype.

Only rarely are the actual disease-causing mutations identified,[7] and progress in the identification of such mutations is hindered by the strong degree of linkage disequilibrium (LD) across the region. Consequently, an MHC haplotype that is associated with a particular disease differs in many (neutral) mutations from other MHC haplotypes, hindering the identification of the actual disease-causing mutations. This warrants a further question, why do haploblocks in the MHC contain so many polymorphisms? These haploblocks are sometimes referred to polymorphic frozen blocks,[14] and the real challenge is explaining why so many polymorphisms can be maintained in the MHC. A unifying theory that can explain the existence of this extraordinary high level of genetic polymorphisms is a better use of Occam's Razor that provides individual explanations for each of the pathologies and disease-causing mutations.

ABC Evolution

Associative Balancing Complex (ABC) evolution proposes that haploblocks with MHC genes ('peri-MHC') accumulate genetic polymorphisms because of their high gene diversity and strong linkage disequilibria. The high gene diversity (heterozygosity), in the MHC implies that recessive deleterious mutations can build up in the peri-MHC because they are rarely expressed as homozygotes. Consequently, purifying selection only rarely operates against those recessive deleterious mutations. Analogous to the theory developed for the self-incompatibility locus evolution in plants,[15] this mutational load is called a 'sheltered load'. In addition, purifying selection is further hampered by functional epistasis between MHC genes.[16] This suppresses the recombination rates in the peri-MHC, which means that mutations can accumulate in a process analogous to Muller's ratchet.[2] Deleterious mutations can accumulate in small, asexual populations because by chance, not a single individual without any mutations manages to reproduce. Without recombination (and ignoring back-mutations), all individuals in future generations will thus carry at least one mutation. This means that the ratchet has clicked, and it can continue to do so resulting in an accumulation of deleterious mutations (i.e., mutational meltdown).

If we think of a haploblock as a small 'asexual' population, Muller's ratchet will have a similar effect on the genetic divergence of the MHC. Functional epistasis reduces genetic recombination between different haploblocks, which means that all haploblocks evolve independently. The genetic divergence between different haploblocks elevates the genetic polymorphism in the MHC significantly above that of the genome-wide average. Much of this genetic polymorphism will be neutral, but some mutations might have a deleterious effect on fitness. These 'bad' mutations can persist as long as they are recessive so that their detrimental effects on fitness are only rarely expressed in homozygous condition. In addition, some deleterious mutations can persist when their resulting change in phenotype is compensated for by an antagonistic mutation in the same haploblock.

ABC Evolution and Trans-Species Polymorphism

ABC evolution reduces the rate of allelic turnover compared to balancing selection and neutrality, because when a new MHC mutant arrives, it will share its sheltered load with its parental haplotype.[1] The novel mutant haplotype expresses its deleterious mutations when in heterozygous condition with a copy of its parental haplotype, which will hinder its invasion into the population. If the new mutant manages to invade, it is likely to replace its parental haplotype. This results in a slow rate of allelic turnover and MHC genealogies that are typical for trans-species evolution with long terminal branches.[1]

MHC haplotype lineages can persist for millions of years with ABC evolution. Distinct MHC haplotype lineages are not lost so that they can continue to track the evolution of the immune-recognition evading parasites. Even if the parasites become locally extinct, this does not jeopardize the existence of the MHC haplotype lineage that conferred the initial resistance (but which now may have become functionally obsolete). Selection against the mutational load of other haplotypes will suppress their frequency once they become common, and this will reduce the risk of loosing haplotypes with such functionally obsolete antigen-binding motifs.

ABC Evolution and HLA Disease Associations

ABC evolution predicts that the deleterious mutations that accumulate in the peri-MHC should remain sheltered from selection when heterozygous, but that they may become expressed in a homozygous state. Consistent with this prediction, many disease-associated haplotypes linked to the HLA are partially recessive.[17] The model also predicts that antagonistic mutations within a haploblock should remain in repulsion phase linkage disequilibrium, and that this suppresses the evolutionary successful recombination rate. Both predictions appear to be consistent with empirical data. Many HLA haplotypes have evolved without showing evidence of recombination over long evolutionary time scales.[18] Furthermore, only recombinant HLA haplotypes appear to show an autoantibody phenotype significantly different from that of control (baseline) individuals.[13] Epistatic gene-gene interactions will thus increase linkage disequilibria across the peri-MHC region, effectively resulting in independently evolving haplotype lineages that can continue to accrue genetic polymorphisms.

ABC Evolution and Genomic Parasites

The high gene diversity and strong linkage disequilibria also make the MHC prone to the onslaught of genomic parasites such as retroviruses and TEs.[19] On average more than 40% of the mammalian genome consists of TEs,[20] and they often reduce the fitness of their host, for example by disrupting coding regions, affecting the recombination rate and gene splicing.[21] However, some TEs have been "domesticated" by their hosts,[22] and they may have gained a regulatory role,[23] or have become part of a gene.[24] Compared to the genome wide average, the MHC appears to have an elevated density of TEs,[23] and some researchers have interpret this as evidence for a potentially constructive evolutionary role of TEs in the MHC.[25] However, given that natural selection is less efficient in the peri-MHC, it is more likely that TEs have accumulated in this genomic region *despite* being mostly deleterious.[19] The irony is that the MHC has evolved as an excellent defense against bacteria, viruses and other parasites, but that this evolutionary innovation made this part of the vertebrate genome susceptible to the assault of genomic parasites such as TEs and retroviruses. This may partly explain why in humans the immune genes that protect us against infectious diseases are associated to more genetic disorders than any part of the our genome.

References

1. van Oosterhout C. A new theory of MHC evolution: beyond selection on the immune genes. Proc R Soc B 2009; 276:657-65.
2. Muller HJ. Some genetic aspects of sex. Am Nat 1932; 66:118-38.
3. Doherty PC, Zinkernagel RM. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature 1975; 256:50-2.
4. Clarke BC, Kirby DRS. Maintenance of histocompatibility polymorphism. Nature 1966; 211:999-1000.
5. Hedrick PW. Pathogen resistance and genetic variation at MHC loci. Evolution 2002; 56:1902-8.
7. Shiina T, Ota M, Shimizu S, Katsuyama Y, Hashimoto N, Takasu M, et al. Rapid evolution of major histocompatibility complex class I genes in primates generates new disease alleles in humans via hitchhiking diversity. Genetics 2006; 173:1555-70.
6. HLA Immunobiology of the Human MHC. Proceedings of the 13th International Histocompatibility Workshop and Congress 2004; IHWG Press, Seatle WA.
8. Klein J, Sato A, Nikolaidis N. MHC TSP, and the origin of species: from immunogenetics to evolutionary genetics. Annu Rev Genet 2007; 41:281-304.
9. Hamilton WD, Zuk M. Heritable true fitness and bright birds: a role for parasites? Science 1982; 218:384-7.
10. Takahata N, Nei M. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics 1990; 124:967-78.
11. Richman AD, Herrera LG, Nash D, Schierup MH. Relative roles of mutation and recombination in generating allelic polymorphism at an MHC class II locus in *Peromyscus maniculatus*. Genet Res Camb 2003; 82:89-99.
12. Charlesworth B. Evolution in Age-Structured Populations. Cambridge University Press 2004; Cambridge.
13. Vandiedonck C, Beaurain G, Giraud M, Hue-Beauvais C, Eymard B, Tranchant C, et al. Pleiotropic effects of the 8.1 HLA haplotype in patients with autoimmune myasthenia gravis and thymus hyperplasia. Proc Natl Acad Sci USA 2004; 101:15464-9.
14. Dawkins R, Leelayuwat C, Gaudieri S, Tay G, Hui J, Cattley S, et al. Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. Immunol Rev 1999; 167:275-304.
15. Stone JL. Sheltered load associated with S-alleles in *Solanum carolinense*. Heredity 2004; 92:335-42.
16. Trowsdale J. The gentle art of gene arrangement: the meaning of gene clusters. Genome Biol 2002; 3:1-5.
17. Gaya DR, Russell RK, Nimmo ER, Satsangi J. New genes in inflammatory bowel disease: lessons from complex diseases? Lancet 367:1271-84.
18. Raymond CK, Kas A, Paddock M, Qiu R, Zhou Y, Subramanian S, et al. Ancient haplotypes of the HLA class II region. Genome Res 2005; 15:1250-7.
19. van Oosterhout C. Transposons in the MHC: the Yin and Yang of the vertebrate immune system. Heredity 2009; Advanced online publication doi:10.1038/hdy.2009.46
20. Margulies EH, NISC Comparative Sequencing Program, Maduro VVB, Thomas PJ, Tomkins JP, Amemiya CT, et al. Comparative sequencing provides insights about the structure and conservation of marsupial and monotreme genomes. Proc Natl Acad Sci USA 2005; 102:3354-9.
21. van de Lagemaat LN, Medstrand P, Mager DL. Multiple effects govern endogenous retrovirus survival patterns in human gene introns. Genome Biol 2006; 7:R86.
22. Miller WJ, McDonald JF, Nouaud D, Anxolabéhère D. Molecular domestication—more than a sporadic episode in evolution. Genetica 1999; 107:197-207.
23. van de Lagemaat LN, Landry JR, Mager DL, Medstrand P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet 2003; 19:530-6.
24. Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. Ann Rev Genet 2007; 41:331-68.
25. Doxiadis GGM, de Groot N, Bontrop RE. Impact of endogenous intronic retroviruses on major histocompatibility complex class II diversity and stability. J Virol 2008; 82:6667-77.